SIGNAL PROCESSING IV: Theories and Applications
J.L. Lacoume, A. Chehikian, N. Martin, and J. Malbos (Editors)
Elsevier Science Publishers B.V. (North-Holland)

INCLUSION OF GRADIENT-BASED SCHEMES IN HIGH-RESOLUTION METHODS

G.Vázquez, F.Vallverdú and M.Amengual

E.T.S.E.Telecomunicació
Apdo. 30.002
08080 Barcelona, SPAIN

ABSTRACT

This paper deals with the inclusion of adaptive schemes in high-resolution methods. Past works refer gradient and conjugated gradient approaches of Pisarenko's spectral line decomposition method for sinusoidal in white noise modelling. The proposal is more general in the objective definition. It is well known that it is possible to generate a complete family of spectral estimators from the Normalized Maximum Likelihood Method, with a structure of power Rayleigh quotien.

In the generation of adaptive versions, the paper proposes the substitution of the conventional mean square filtering error by quadratic objectives built as inner products of the coefficient error vector of the estimator filter.

1.-INTRODUCTION

The main goal of this paper is the inclusion of adaptive methods in high-resolution spectral estimation algorithms. High-resolution refers to spectral analysis methods that try to estimate a number of sinusoidal signals or sources in noise. Very often, we have short data segments, and thus, use of conventional methods lead to very low resolution estimations. If there is any additional available information of the process we are analyzing, it can be included in the method with the use of certain models of the process for improving the estimation performances.

Most of the part of such kind of algorithms are based on the Singular Value Decomposition (S.V.D.) of the data autocorrelation matrix /3/, /4/, and thus with a hard computation cost. This effort is reduced in algorithms with a partial decomposition, as for Pisarenko, in which minimum eigenvalue and its combined eigenvector are only necessary.

As an attractive alternative to these methods the Generalized Power Maximum Likelihood Method /2/ can be used , as a consequence of the previous Normalized Maximum Likelihood Method (N.M.L.M.) /2/. In such a way Blackman-Tuckey, Maximum Likelihood (or Capon), and the N.M.L. Methods were included in the generalized estimator as particular cases , and with the possibility for extending it to high-resolution diagrams. The same authors introduced the idea for seeing these methods as a bank of filter spectral estimation approach, and giving a unified point of view. Fortunately, this interpretation has been extended to the high-resolution algorithms /5/,/7/. It is possible to see the Generalized Power M.L. Method as a consequence of the Normalized M.L. Method previous the inclusion of some linear transformations over the data. These pre-processing linear transformations will let us to generate adaptive schemes for improving resolution.

The inclusion of linear transformations previous to adaptation make fall in crisis some of the ideas used in the generation of adaptive gradient-based algorithms. To avoid these difficulties we will introduce a generalized cost function as objective to minimize, substituting the conventional mean square filtering error function by inner products of the coefficient error vector of the estimator filter.

2.-BACKGROUND

Given a random, stationary process x(n), the autocorrelation sequence is defined as the second order statistics:

$$r(m) = E(x(n+m)\, x^{*}(n)) \qquad (1)$$

where, in general, the expectation is substituted by a certain estimation over the available data record. The power spectrum follows from the Fourier Transform of the autocorrelation function (1):

$$S(\omega) = \sum_{m} r(m) \exp(-jm\omega) \qquad (2)$$

This work is supported by the Spanish Research Council under grand PR83-2906-C03-02

If the process x(n) has been degraded by additive noise, the spectrum will show components due to signal, but overlapped the contribution of noise. Under some conditions it is possible to separate both components.

If we define the data vector X(n) as follows:

$$X(n) = (x(n),\ x(n-1),..,x(n-N+1)) \qquad (3)$$

for a matrix order N, the autocorrelation matrix is introduced by:

$$R = E(X(n)\ X^H(n)) \qquad (4)$$

This matrix is assumed to be Toeplitz, positive definite and hermitian, thus, it is possible to decompose it in a modal way:

$$R = Q\ \Lambda\ Q^H \qquad (5)$$

where Q is the orthonormal eigenvector matrix and the diagonal eigenvalue matrix:

$$Q = (Q_1,\ Q_2,\ \ldots,\ Q_N) \qquad (6)$$

$$\Lambda = (\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_N) \qquad (7)$$

with a convenient order:

$$\lambda_1 > \lambda_2 > \ldots\ldots\ldots \geq \lambda_N > 0$$

Assuming that process x(n) has been generated by a combination of M complex sinusoidal signals of arbitrary amplitudes and phases in white noise of λ_{min} mean power, the autocorrelation matrix will present the (N-M) least significant eigenvalues all equal to the value λ_{min}:

$$\lambda_{M+1} = \ldots = \lambda_N = \lambda_{min} \qquad (8)$$

The study of rank for matrix $(R - \lambda_{min} I)$ stablishes an important decomposition for the space in the signal subspace generated by the M eigenvectors combined to the M most significant eigenvalues, and the noise subspace that is generated by the (N-M) remaining eigenvectors.

The constraint that noise must be white may be relaxed and the results remain unchanged if the modal decomposition is made in the metric of the auto- correlation matrix for the noise.

This decomposition lets us a better description of the process, and the different spectral estimation methods differs in how the information is used to build it, choosing the noise or the signal eigenvalues and eigenvectors.

The orthogonality of eigenvectors ensures that the frequency response of the noise eigenvectors will present a null response to the frequencies of the pure sinusoids. This property is specially important in the development of high-resolution algorithms. Besides, matrix R is hermitian and it lets us to express any power of matrix R in a simple way:

$$R^k = Q\ \Lambda^k\ Q^H \qquad (9)$$

Thus, for high values of k, the set of eigenvector remains unchanged and only the eigenvalue spread is emphasized ,and through an adequated value of k it is possible to choose the signal subspace(k>0) or the noise subspace (k<0).

Maximum Likelihood Method is an example of a estimator built with negative powers of the autocorrelation matrix, and the resolution is much better than for Blackman-Tuckey. The M.L. Method is not really a power density estimator. It corresponds to a power measure, and it must be normalized by the bandwidth of the filter we are using in the measurement. This point suggested the Normalized M.L. Method, with a Rayleigh quotient structure.

From the expression for the M.L.M. , Capon /1/ proposed the following generalized family estimators, where he introduced a power function of the autocorrelation matrix:

$$S^q_p(\omega) = (S^H R^q S)^{1/q} \qquad (10)$$

that leads to B.T. for q=1 and to M.L.M. for q=-1. For the rest of q values, relation (10) losses the meaning of spectrum estimator and becomes 'frequency detector' in the signal subspace for q>0 or in the noise subspace for q<0.

The Power Generalized Maximum Likelihood Method (Q.M.L.M.) is given by the following family of expressions:

$$S^q_{ML}(\omega_o) = \frac{S^H R^{-q+1} S}{S^H R^{-q} S} \qquad (11)$$

that supplies B-T for q=0, M.L.M. for q=1 and N.M.L.M. for q=2. The Rayleigh quotient structure for (11) ensures that for all q:

$$S^{q-1}_{ML}(\omega) \geq S^q_{ML}(\omega)\ ;\ \forall\ q \geq 0 \qquad (12)$$

and resolution improves with value q.

For a general treatment, let's consider the transformed data vector:

$$\tilde{X}(n) = T^H X(n) \qquad (13)$$

with an autocorrelation matrix given by:

$$B = T^H E(X(n)\ X^H(n))\ T = T^H R\ T \qquad (14)$$

The output power for the filter with coefficients W and for the transformed data is expressed by:

$$P_o = W^H B\ W \qquad (15)$$

Filter N.M.L.M. for the new data becomes:

$$W' = \frac{B^{-1} S}{S^H B^{-1} S} \qquad (16)$$

and the bandwidth of the filter is given by the norm of (16). Actually, it is not correct to strengthen that the Q.M.L.M.(11) is the consequence of transforming the data and then introducing a certain estimator. It is used over the original data. Then, the output power with coefficient vector (16) can be expressed as:

$$P_0 = W'^H R W' \qquad (17)$$

The normalized power supplies the final estimator:

$$S_T(\omega) = \frac{S^H B^{-1} R B^{-1}}{S^H B^{-2} S} \qquad (18)$$

It is possible to find an adequate transformation T (13) such that:

$$B = T^H R T = R^{q/2} \qquad (19)$$

Including (19) in (18), we obtain the Q.M.L.M. expression (11) for:

$$T = R^{(q-1)/2} \qquad (20)$$

This development suggests the possibility of including that kind of linear transformations ((13) and (20)) in linear adaptive filtering and prediction. The main consequence will be the generation of new quadratic objectives for obtaining the adaptive schemes.

3.- FILTERING OF TRANSFORMED DATA

In this section we will study the influence of autocorrelation-matrix-based linear transforms in quadratic objectives. For a reference sample d(n) we try to estimate it as a linear combination of the transformed data samples $\tilde{X}(n)$ (13):

$$\hat{d}(n) = W^H \tilde{X}(n) \qquad (21)$$

where:

$$\tilde{X}(n) = R^k X(n) \qquad (22)$$

The estimation error is given by:

$$e(n) = d(n) - \hat{d}(n) \qquad (23)$$

and the mean square error (m.s.e.) for (23):

$$\xi^2_k = E(|e(n)|^2) =$$
$$W^H R^{2k+1} W - W^H R^k P - P^H R^k W + r_d(0) \qquad (24)$$

The optimal solution that minimizes the error (24) becomes:

$$W^* = R^{-(k+1)} P \qquad (25)$$

Including (25) into (24), we get the minimum quadratic mean error:

$$\xi^2_{k\,min} = r_d(0) - P^H R^{-1} P \qquad (26)$$

that is k value independent, and thus, solution (25) lays on a constant error plane, and common to the Wiener solution. The m.s.e. can be expressed as an inner product of the weight error vector:

$$\xi^2_k = (W - W^*)^H R^{2k+1} (W - W^*) + \xi^2_{min} \qquad (27)$$

where W^* denotes the optimum solution (25). Previous expression shows that the m.m.s.e. is a distance between filter weights and optimum ones, that is, as a measure of how the filter weight vector fits the optimum solution. As R is assumed positive, expression (27) is well defined and with a unique solution. It is possible to design other functions of the weight error vector and with the same optimum. For instance, we can generalize the objective by the following inner-product:

$$\xi^2_{k,m} = (W - W^*)^H R^m (W - W^*) + \xi^2_{min} \qquad (28)$$

The interest of (28) is that the eigenvalue spread of the weighting matrix is modified by the value m, but the solution remains unchanged (25). The m.s.e. objective (27) is only a particular case of (28), that is, for m = 2k + 1:

4.- FILTERING AND WEIGHT-VECTOR ERROR GRADIENT APPROACHES

The simplest scheme approaching the solution is to move an initial guest W(0) for the coefficients in the direction of the gradient for the m.s.e. objective (27):

$$\nabla \xi_{2k,2k+1}(n) = R^{2k+1} (W(n) - W^*) \qquad (29)$$

Taking instantaneous expressions in (29), the simplest filtering gradient estimate is the following:

$$\nabla \xi_{2k,2k+1}(n) = - e^+(n) \tilde{X}(n) \qquad (30a)$$

with the 'a priori' error defined by:

$$e(n) = d(n) - W^H(n) \tilde{X}(n) \qquad (30b)$$

The gradient actualization diagram is given by:

$$W(n+1) = W(n) + \mu e^+(n) \tilde{x}(n) \qquad (31)$$

that corresponds to the well known L.M.S. algorithm. The convergence of the method is ensured in mean for any 'step-size' such that:

$$0 \leq \mu < 2/(\lambda_{max})^{2k+1} \qquad (32)$$

The convergence time for the m.s.e.(29) is proved to be:

$$T_{min} > \frac{1}{4} \left(\lambda_{max}/\lambda_{min}\right)^{2k+1} \qquad (33)$$

Thus, for ill conditioned matrix or for high values of k, the convergence time could be unacceptable.

Taking into account that the generalized objective (28) suplies the same optimal weight vector, we can try to develop gradient diagrams from it. The generalized gradient is given by:

$$\nabla \xi_{2k,m} = R^{m} (W - W^{*}) \qquad (34)$$

The most attractive case leads for m=0. In that case, the method exhibits a data independent performance, that is, without sensitivity to the eigenvalue spread as for the rest m values:

$$W(n+1) = W(n) + \mu R^{-(k+1)} e^{+}(n) X(n) \qquad (35)$$

This actualization alternative corresponds to the classical R.L.S. method, and it is possible to give it an exact recursive diagram. The main handicap is the computation of $R^{-(k+1)}$ that needs the use of the matrix inversion lemma for some topics as for instance in arrays, but can be optimized with fast versions for spectral estimation and transversal filtering (Fast Kalman, FAEST, FTF) /5/.

The tradeoff between performance and complexity is obtained for a third case in (34). Taking m=k+1 in (34) and instantaneous estimates, the coefficients are updated by:

$$W(n+1) = W(n) + \mu e^{+}(n) X(n) \qquad (36)$$

Stability is ensured for:

$$0 \leqslant \mu < 2/\left(\lambda_{max}\right)^{k+1} \qquad (37)$$

with a convergence time much better that for L.M.S. (31)(33) and with the same complexity:

$$T_{min} > \frac{1}{4} \left(\lambda_{max}/\lambda_{min}\right)^{k+1} \qquad (38)$$

Misadjustment is also improved by the same ratio.

The performances of the algoritms can be described by the 'demerit factor D' obtained as the product of the convergence time and the misadjustment error. For L.M.S. algorithm (30b)(31) we get the approximation:

$$D_{LMS} = \frac{N}{4} \sum_{i} \left(\lambda_{i}/\lambda_{min}\right)^{2k+1} \qquad (39)$$

Thus, this quality mesure shows that the performance decrease with value 'k'. The equivalent parameter for the modified version of the L.M.S. (30b)(36) as follows:

$$D_{MOD} = \frac{N}{4} \sum_{i} \left(\lambda_{i}/\lambda_{min}\right)^{k+1} \qquad (40)$$

The lowest bound for the demerit factor is found the Newton diagram developed in (30b)(35), and now is given by:

$$D_{N} = \frac{N}{4} \qquad (41)$$

and, it is independent of the data eigenval distribution and of 'k'.

5.- REFERENCES

/1/ V.F. PISARENKO. 'On the Estimation of Spect by Means of Non-Linear Functions of the Covarian Matrix'. Geophys. J.R. Astr. Soc., Vol. 28, 197 pp.511-531.

/2/ M.A. LAGUNAS ET AL. 'An Improved Maxim Likelihood Method for Power Spectral Densi Estimation'. IEEE Trans. ASSP, Vol. ASSP-32, No. Feb. 1984.

/3/ R.O. SCHMIDT. 'Multiple Emitter Location a Signal Parameter Estimation'. Trans. Antennas a Propagations, Vol. AP-34, No.3, March 1986.

/4/ J.M. SPEISER, ET AL. 'Signal Processi Computations Using the Generalized Singular Val Decomposition'. SPIE Proceedings, Vol.495, S Diego, CA, 21-22, Aug. 1984.

/5/ J.M. CIOFFI, ET AL. 'Fast, Recursive Lea Squares Transversal Filters for Adapti Filtering'. IEEE Trans. ASSP, Vol. ASSP-32, No. April 1984.

/6/ M.A. LAGUNAS. 'Power Functions in Spectr Estimation'. To be published in ICASSP-88. N York, U.S.A.

/7/ M.A. LAGUNAS and M. CABRERA. 'Rayle Estimates: Performing like S.V.D.'. EUSIPCO-88.